Twelve Cities

Jonathan Auerbach and Rob Trangucci

12/1/2016

Summary of Analysis

We investigate whether American cities can expect to achieve a meaningful reduction in pedestrian deaths by lowering the posted speed. We present our work in three sections. First we motivate the problem and provide a description of the dataset. Second we fit a log-linear model and compare likely sources of confounding with an analysis of variance. In the last section we demonstrate a sample use case. We evaluate the decision to lower many of New York City's posted speed limits from 30 mph to 25 mph. In our evaluation, we assume speed limits are assigned ignorably given measured covariates, and we calculate the number of lives saved by estimating the causal effect of lowering the speed limit on New York City roads from 30 mph to 25 mph on 25 mph roads. Our estimated causal effect is much smaller than the traditional before-after analysis favored by policy analysts.

I. Introduction

The Policy Question

Over the last few years, American cities have started to establish aggressive countermeasures that force road users to make safer decisions. These policies are collectively known as Vision Zero and have a stated goal of creating a road system with zero traffic fatalities.

In theory, Vision Zero is a comprehensive road investment strategy that prioritizes safety over vehicle mobility by anticipating human error and then slowing vehicles to the safest possible travel speed. Citywide road redesign and increased levels of enforcement and outreach are integral to a Vision Zero strategy (Tingvall and Haworth, 1999). These approaches are effective because they confine drivers, considerably reducing the chance that human error will result in a fatality. Yet such changes are costly to implement and challenging to sustain across an entire city. In practice, policymakers simply mandate that vehicles reduce their travel speed by lowering the posted speed limit below the speed for which the road was originally designed.

Lowering the speed limit is politically expedient since speed limit changes can be implemented immediately and at relatively little cost, and they can be sustained across an entire city at no additional expense. But lowering a speed limit without improvements to road design, enforcement and outreach may do little to reduce fatalities if drivers feel little pressure to comply with the lower limit (Leaf and Preusser, 1999). In fact, the National Highway Safety Traffic Administration (NHSTA) rates the countermeasure "reduce and enforce speed limits" three out of five stars for improving safety because research indicates that actual speed is reduced by only a fraction of the lowered speed limit (Goodwin et al., 2015).

Twelve major American cities have officially set a Vision Zero strategy in response to vocal advocates: Chicago, San Francisco, New York City, Boston, Los Angeles, Austin, Portland, Seattle, San Jose, San Diego, Washington D.C. and Denver. The majority of these strategies include adjustments to many or all of the posted speed limits. For example, New York City lowered the default citywide speed limit from 30 to 25 mph in 2015. San Francisco uniformly set the speed limit around schools to 15 mph, and is considering a default citywide speed limit of 20 mph.

The policy question is twofold. The first is descriptive: what aspects of roads drive pedestrian fatalities? The second is prescriptive: can a speed limit adjustments meaningfully reduced the number of pedestrian deaths?

The Data

The dataset used for this analysis contains every pedestrian death in the twelve major American cities with Vision Zero policies between 2010 and 2015 for which there was no missing covariate information. The unit of our analysis, however, is the immediate region containing the road segment on which each death occurred. We consider the dataset to be a random sample of these regions, and we have constructed sampling weights that estimate the probability the region would be included in the sample.

The dataset has the following eight qualitative variables describing each road region: the weather and surface condition (COND), the city (CITY), the year (YEAR), the posted speed limit (SLIM), the presence of different signs or signals (SIGN), the time and lighting (LGHT), the physical road characteristics or built environment (BLTE) and the annual average traffic density (TFFC). The dataset has two quantitative variables: the relative number of pedestrians exposed to the road (EXPR) and the probability of the road being included in the sample (WGHT). More information on the dataset can be found in the Appendix.

The first five observations of the dataset are displayed in Table 1 below. Table 2 shows summary statistics for the most common qualitative variable levels. Figure 1.1 plots the histogram of observations by city. Figure 1.2 plots these cities by the most common speed limits, and Figure 2 plots them by both the most common speed limit and the last four years of the dataset.

```
kable(head(crash_data),caption="Example Observations of Dataset")
```

Table 1: Example Observations of Dataset

COND	CITY	YEAR	SLIM	SIGN	LGHT	BLTE	TFFC	EXPR	WGHT
2	3	1	11	1	1	13	1	1300.3571	0.0465116
2	3	1	6	1	25	14	1	196.0476	0.3594917
2	8	1	8	1	21	56	4	398.0952	0.3982802
2	8	1	8	1	21	22	1	497.3810	0.4145551
2	10	1	9	1	12	42	1	458.0357	0.4280646
24	8	1	8	1	12	54	4	317.7679	0.4565141

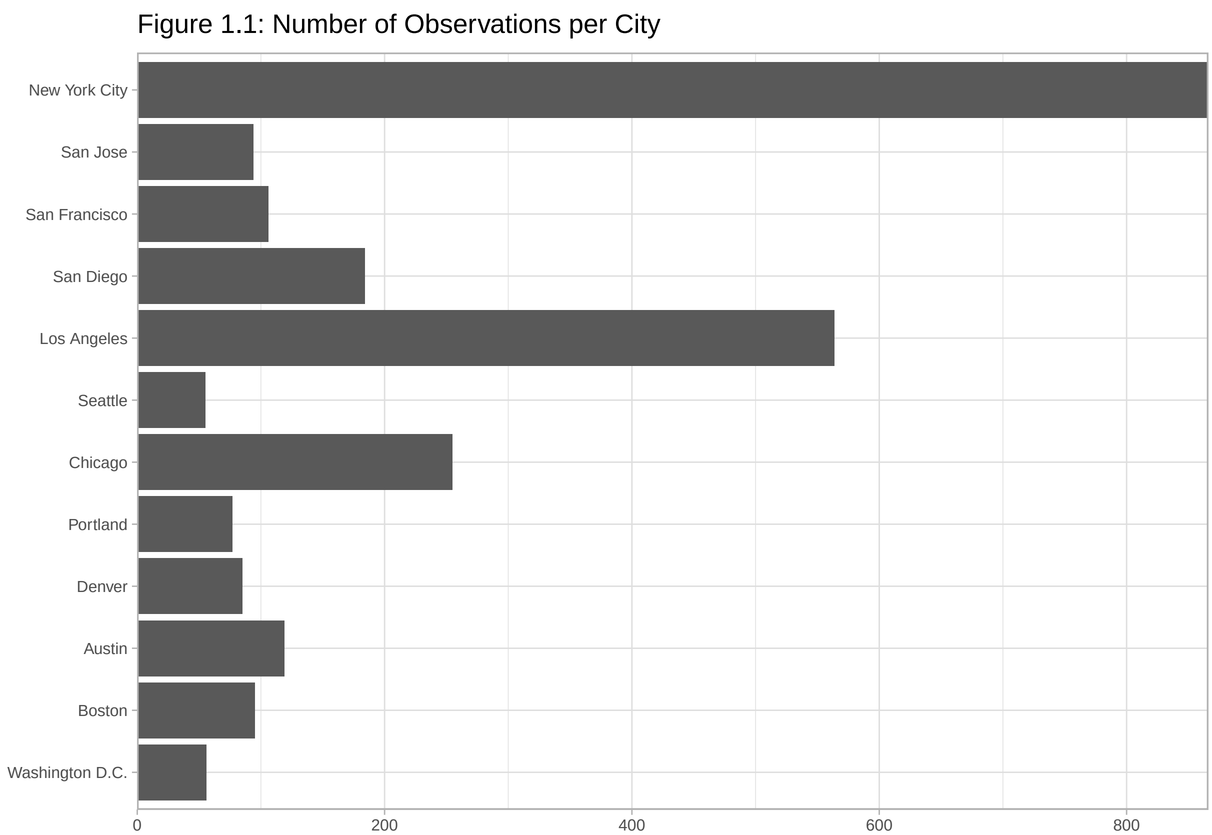
```
kable(summary(apply(crash_data[,1:8],2,factor)),caption="Summary of Dataset")
```

Table 2: Summary of Dataset

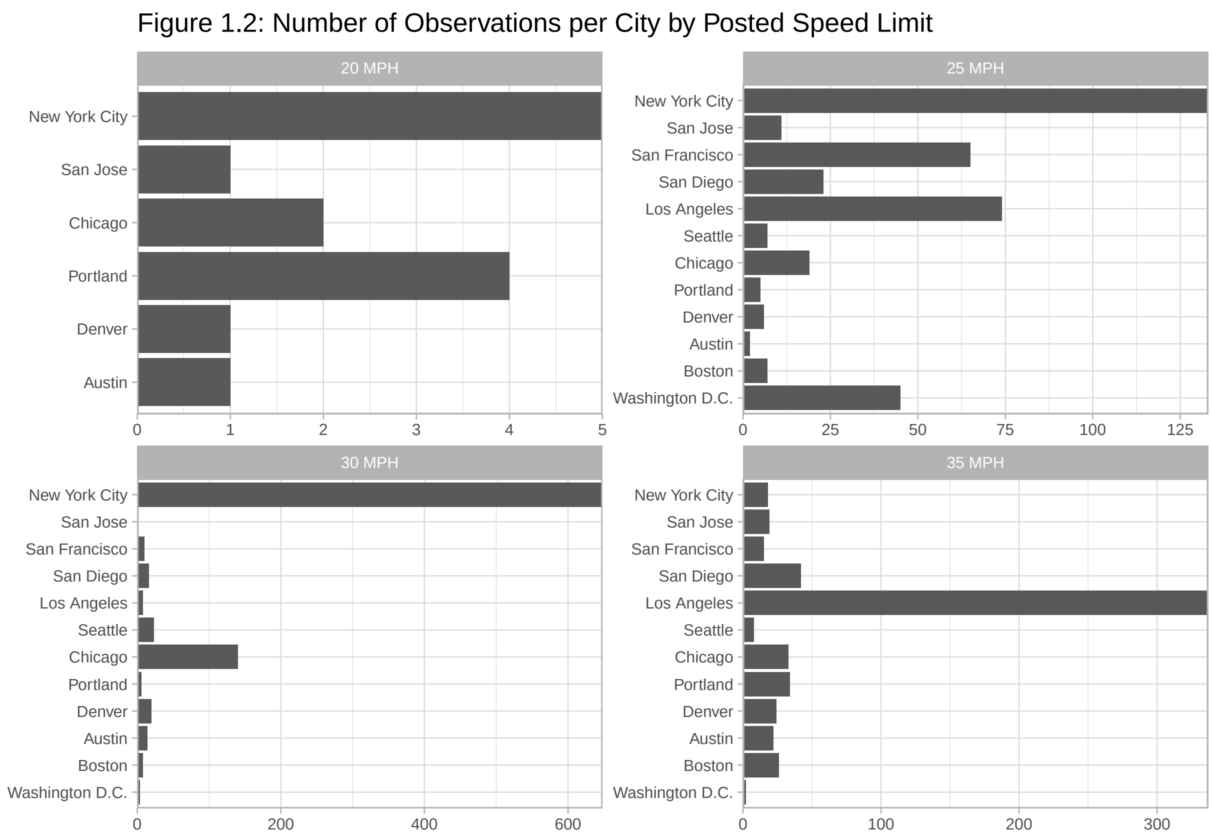
COND	CITY	YEAR	SLIM	SIGN	LGHT	BLTE	TFFC
2 :1976	12 :866	1:409	7 :898	1 :1578	8 :386	16 : 358	1:546
23 : 232	8 :564	2:398	8 :580	4 : 476	21 :357	56 : 259	2:830
9 : 198	6 :255	3:443	6 :397	3 : 340	29 :227	14 : 243	3:640
24 : 56	9 :184	4:455	10 :170	6 : 78	1 :223	54 : 192	4:540
3 : 24	3 :119	5:406	9 :136	9 : 24	34 :221	41 : 77	NA
18 : 16	10 :106	6:445	14 :135	2 : 19	25 :161	79 : 63	NA
(Other): 54	(Other):462	NA	(Other):240	(Other): 41	(Other):981	(Other):1364	NA

```
ggplot(crash_data) +
  theme_light(8) +
  aes(factor(CITY,labels=c(
    "Washington D.C.","Boston","Austin","Denver","Portland","Chicago",
    "Seattle","Los Angeles","San Diego","San Francisco","San Jose",
```

```
    "New York City"))) +
  geom_histogram(stat="count") +
  coord_flip() +
  scale_y_continuous(expand = c(0, 0)) +
  labs(x="",y="",
   title="Figure 1.1: Number of Observations per City")
```

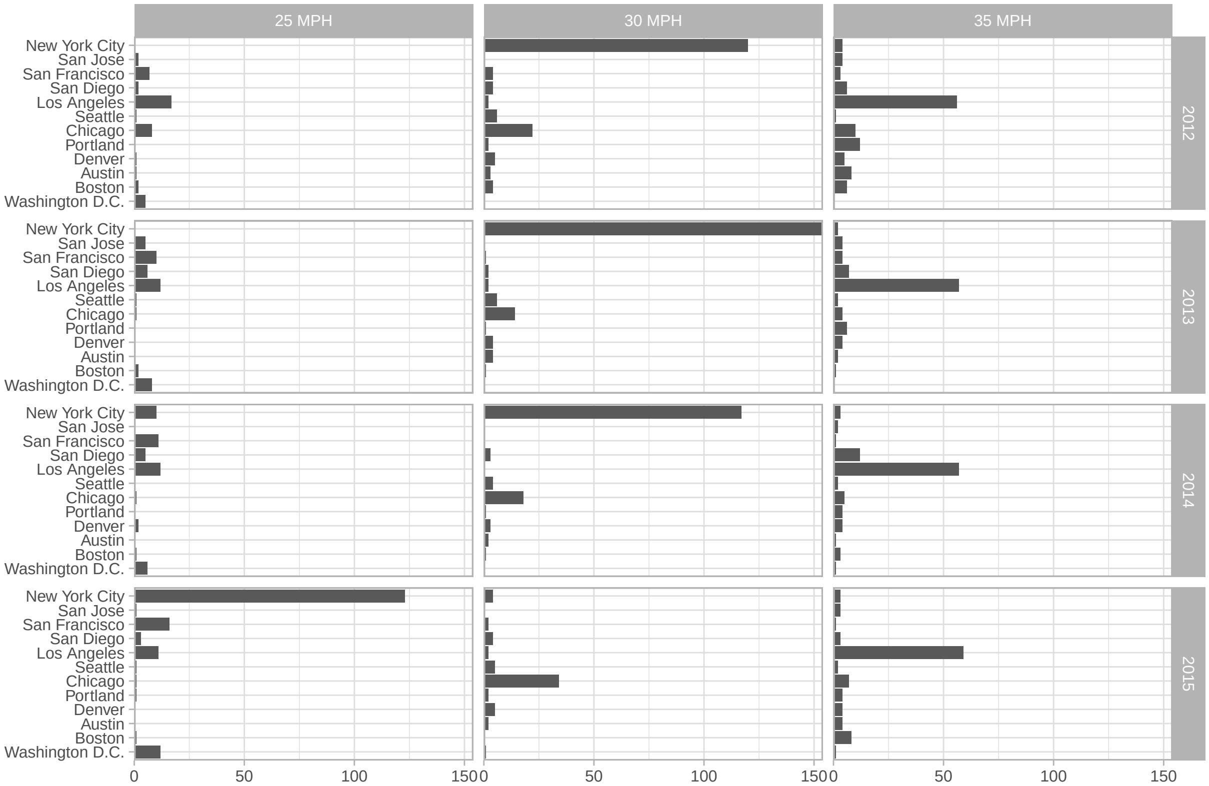



```
ggplot(crash_data[crash_data$SLIM %in% 5:8,]) +
  theme_light(8) +
  aes(factor(CITY,labels=c(
    "Washington D.C.","Boston","Austin","Denver","Portland","Chicago",
    "Seattle","Los Angeles","San Diego","San Francisco","San Jose",
    "New York City"))) +
  geom_histogram(stat="count") +
  coord_flip() +
  labs(x="",y="") +
  facet_wrap(~factor(SLIM,labels = c("20 MPH","25 MPH","30 MPH","35 MPH")),
             scales="free") +
  scale_y_continuous(expand = c(0, 0)) +
  labs(x="",y="",
   title="Figure 1.2: Number of Observations per City by Posted Speed Limit")
```




```
ggplot(crash_data[crash_data$SLIM %in% 6:8 &
                  crash_data$YEAR %in% 3:6,]) +
  theme_light(8) +
  aes(factor(CITY,labels=c(
    "Washington D.C.","Boston","Austin","Denver","Portland","Chicago",
    "Seattle","Los Angeles","San Diego","San Francisco","San Jose",
    "New York City"))) +
  geom_histogram(stat="count") +
  coord_flip() +
  labs(x="",y="") +
  facet_grid(factor(YEAR,labels = c("2012","2013","2014","2015"))~
             factor(SLIM,labels = c("25 MPH","30 MPH","35 MPH"))) +
  scale_y_continuous(expand = c(0, 0)) +
  labs(x="",y="",
       title="Figure 1.3: Number of Observations per City by Posted Speed Limit and Year")
```

Figure 1.3: Number of Observations per City by Posted Speed Limit and Year



The 2015 reduction in New York City's default speed limit is obvious in Figure 1.3. For this analysis, we will construct a model for the years 2010-2014 and calculate the causal effect of the 2015 policy. We will also compare our results with the classic before and after analysis favored by policy makers and transportation researchers (Hauer 2005; Hauer 2015).

II. Model and ANOVA

The task of modeling the relationship between the posted speed limit and pedestrian deaths is complicated by a large number of qualitative, confounding variables. The data constitute a high dimensional contigency table where each cell is a low, non-zero count. We propose a non-nested, hierarchical, log-linear model where observations are partially pooled across strata. Causal estimates can then be calculated by averaging over the relevant strata (Gelman and Hill 2007; Gelman et al. 2013). The validity of these estimates depends on a variety of assumptions. Here we highlight three major assumptions.

Assumptions

We make the stable unit treatment value assumption which means we assume no interference between the units of analysis, the road regions. This assumption would be violated if the death rate in any road region depended on the posted speed limit of a neighboring roads region. While such interferences are possible, we anticipate meaningful interactions between roads to be rare since the factors leading to collisions are generally contained within a single road region.

We assume the assignment of posted speed limits is ignorable given the observed covariates. This means that roads in similar conditions (in a particular city, near a school, up a hill, etc.) will be equally likely to be given a particular speed limit as any other. In addition, we assume every road has a positive probability of receiving any speed limit.

We also rely on the following Bayesian model to predict the counterfactual outcomes. In this case, the treatment variable and all of the covariates are qualitative, and we use a zero truncated log-linear model to predict the number of deaths within each strata. We partially pool across strata with similar levels of qualitative variable. This coincides with the prior belief that roads with similar weather conditions or number of lanes, for example, will yield similar fatality rates.

Let y_{ij} denote the i^{th} death in the j^{th} covariate strata. The joint probability distribution can be writen:

$$\begin{aligned}
\epsilon &\sim \text{ Normal}(0, \sigma_\epsilon) \\
\alpha_i^{\cdot} &\sim \text{ Normal}(0, \sigma_i) \\
\bar{y}_{\cdot j} &\sim \text{Poisson}^+(\exp(\mu + \alpha_1^{\text{SLIM}} + \alpha_2^{\text{CITY}} + \alpha_3^{\text{YEAR}} + \alpha_4^{\text{COND}} + \alpha_5^{\text{SIGN}} + \alpha_6^{\text{LGHT}} + \alpha_7^{\text{BLTE}} + \alpha_8^{\text{TFFC}} + \epsilon_j + \beta \cdot \log(\overline{\text{EXPR}})_{\cdot j}))
\end{aligned}$$

We also put weakly informative priors on μ, β and the σ's. However, these priors appear to have little influence on the quantities of interest.

Model Fit

We run the above model with four chains for 500 iterations yielding 1000 posterior samples after warm-up. In the generated quantities block, we calculate the finite sample standard deviations of each qualitative variable to identify major sources of variation (Gelman and Hill 2007). We also predict the counterfactual outcomes for each road at 25 and 30 mph. While we use all of the observations to fit the model parameters, we will only use predictions for New York City roads in 2015 with 25 and 30 mph speed limits for evaluating New York City roads in 2015 with 25 mph speed limits.

```
G <- 8
J <- apply(crash_data[,1:8],2,function(x) length(unique(x)))
stan_dat <- aggregate(cbind(1,EXPR)~.,crash_data,sum)
stan_dat <- stan_dat[order(stan_dat$YEAR),]
stan_dat_list <- with(stan_dat,
                      list(N_train = which.max(stan_dat$YEAR)-1,
                           N = nrow(stan_dat),
                           J = J,
                           G = G,
                           COND = COND,
                           CITY = CITY,
                           YEAR = YEAR,
                           SLIM = SLIM,
                           SIGN = SIGN,
                           LGHT = LGHT,
                           BLTE = BLTE,
                           TFFC = TFFC,
                           count = V1,
                           EXPR = EXPR))

model <- stan_model(file = 'model.stan')
```

```
## In file included from filed85707604e4.cpp:8:
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
```

```
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/include/stan/math/rev/core
##     static void set_zero_all_adjoints() {
##                 ^
## In file included from filed85707604e4.cpp:8:
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/include/stan/math/rev/core
##     static void set_zero_all_adjoints_nested() {
##                 ^
## In file included from filed85707604e4.cpp:8:
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/include/stan/math/prim/mat
##       size_t fft_next_good_size(size_t N) {
##              ^
## In file included from filed85707604e4.cpp:8:
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/StanHeaders/incl
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost
## In file included from /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost/multi_array/concept_c
##       typedef typename Array::index_range index_range;
##                                           ^
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost/multi_array/concept_c
##       typedef typename Array::index index;
##                                     ^
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost/multi_array/concept_c
##       typedef typename Array::index_range index_range;
##                                           ^
## /Library/Frameworks/R.framework/Versions/3.3/Resources/library/BH/include/boost/multi_array/concept_c
##       typedef typename Array::index index;
##                                     ^
## 7 warnings generated.
```

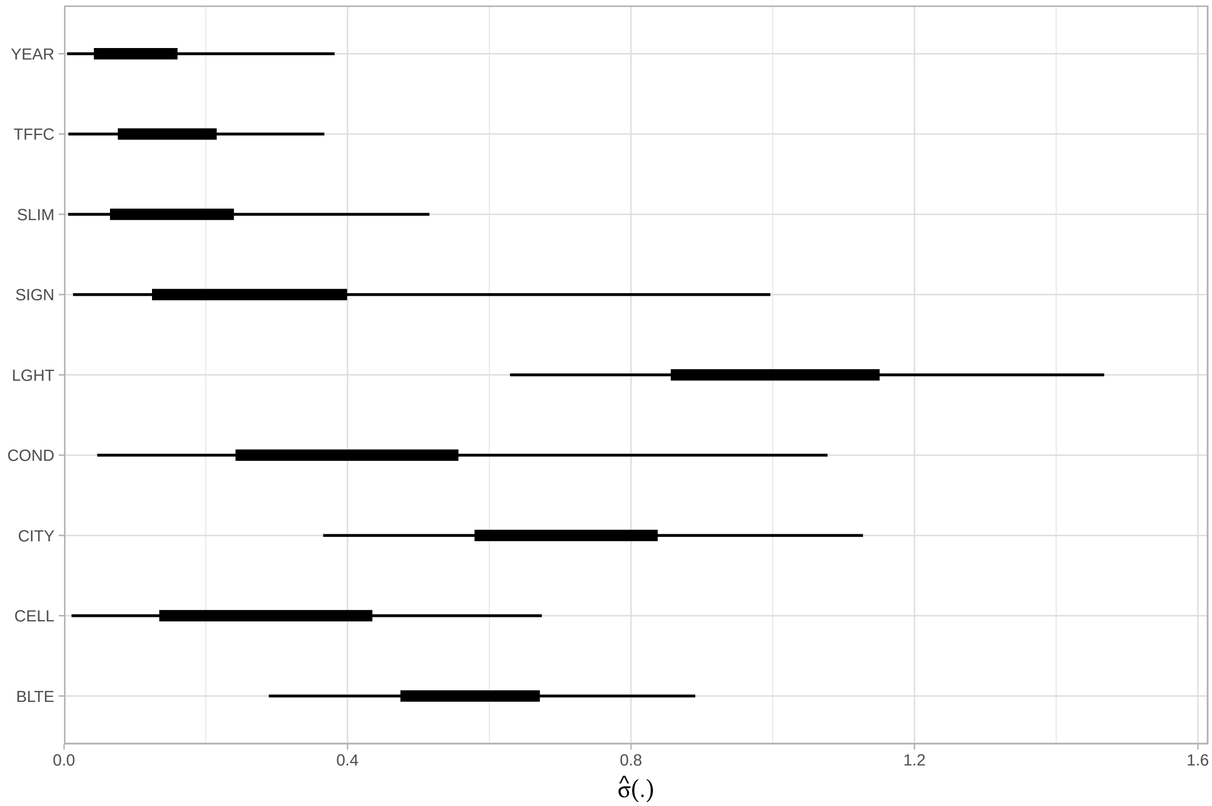
```
fit <- sampling(model, data = stan_dat_list, iter = 500,
                chains = 4, refresh = 10, cores = 4)
```

Analysis of Variance

Figures 2 and 3 display the Analysis of Variance. Figure 2 shows the 50 and 95% quantiles of the finite sample standard deviation. It demonstrates that time of day and lighting (LGHT) explain a significant portion of the pedestrian death rate. This makes sense as this variable is the only one to capture the varying use of roads (i.e. commuting, the bar scene, tourism etc.). City and the built environment (BLTE) also explain much variation in pedestrian deaths. Surprisingly, there is little variation of deaths between cells (CELL), years and levels of traffic (TFFC) suggesting that, through time of day and lighting, city and the built environment, we have adjusted for potential sources of confounding.

```r
btwn <- c("COND_sd","CITY_sd","YEAR_sd",
          "SLIM_sd","SIGN_sd","LGHT_sd",
          "BLTE_sd","TFFC_sd","cell_sd")
coefs <- data.frame(extract(fit,pars=btwn))
coef_ggplot <- data.frame(coef_mean=apply(coefs,2,mean),btwn = btwn)
coef_ggplot$upper50 <- apply(coefs,2,quantile,probs=.75)
coef_ggplot$lower50 <- apply(coefs,2,quantile,probs=.25)
coef_ggplot$upper95 <- apply(coefs,2,quantile,probs=.975)
coef_ggplot$lower95 <- apply(coefs,2,quantile,probs=.025)

ggplot(coef_ggplot, aes(toupper(substr(btwn,1,4)), coef_mean)) +
  theme_light(8) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  coord_flip() +
  labs(y=expression(hat(sigma)(.)),x="",
       title="Figure 2: Analysis of Variance, Between") +
  scale_y_continuous(limits=c(0,1.1*max(coef_ggplot$upper95)),
                     expand = c(0, 0))
```

Figure 2: Analysis of Variance, Between

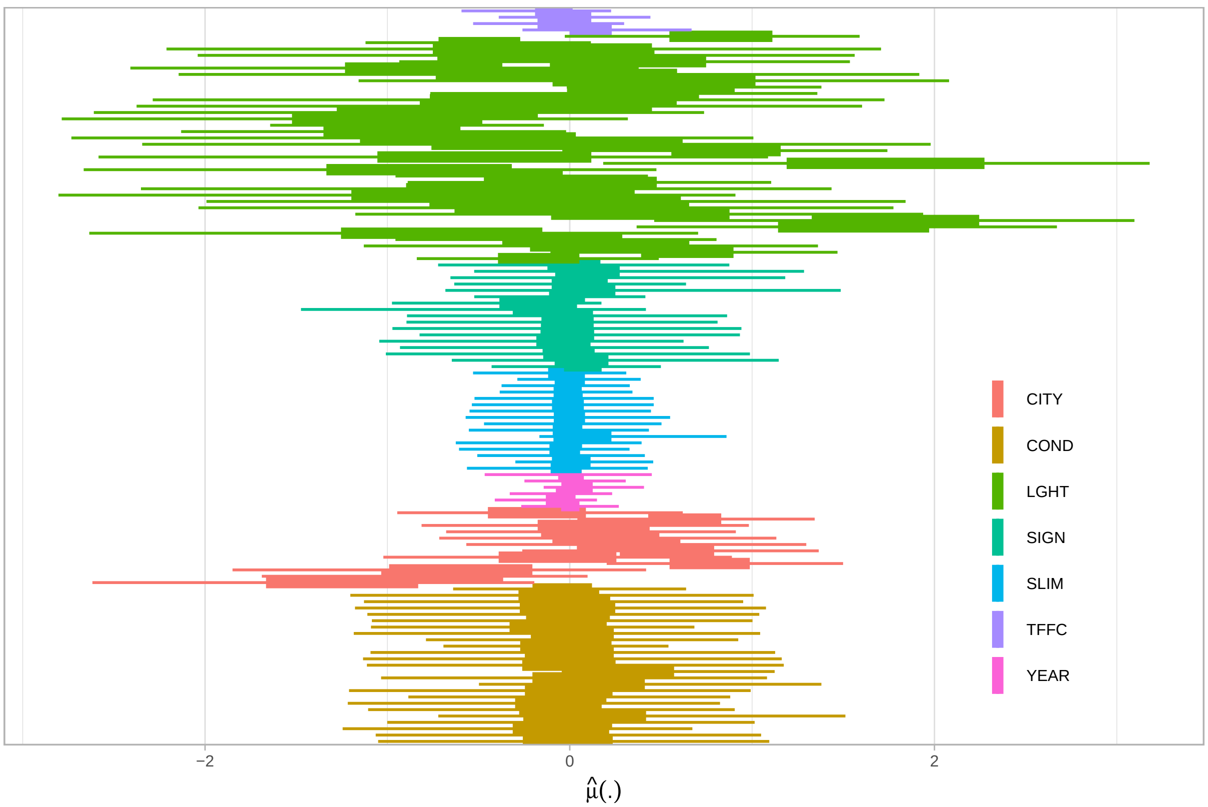


The following four figures demonstrate some of the individual drivers of pedestrian deaths. Figures 3.1 and 3.2 show the 50 and 95% quantiles of the mean effects. Figures 3.3 and 3.4 show just the city effects and speed limit effects respectively. Figure 3.4 shows that pedestrians are, in general, at similar risk of death at 30 as compared with 25 MPH roads. Of course, this is not a comparison of the speed limit effects within the same strata. We make those comparisons in the following section.

```
wthn <- c("COND_e","CITY_e","YEAR_e","SLIM_e",
          "SIGN_e","LGHT_e","BLTE_e","TFFC_e")
coefs <- extract(fit,pars=wthn)
coef_ggplot <- data.frame(wthn = character(),
                          numb = character(),
                          coef_mean=numeric(),
                          upper50 = numeric(),
                          lower50 = numeric(),
                          upper95 = numeric(),
                          lower95 = numeric())
for(var in seq_along(coefs)){
  coefs_temp <- coefs[[var]]
  coef_ggplot_temp <- data.frame(wthn = wthn[var],
                                 numb = paste0(wthn[var],1:ncol(coefs_temp)),
                                 coef_mean=apply(coefs_temp,2,mean))
  coef_ggplot_temp$upper50 <- apply(coefs_temp,2,quantile,probs=.75)
  coef_ggplot_temp$lower50 <- apply(coefs_temp,2,quantile,probs=.25)
  coef_ggplot_temp$upper95 <- apply(coefs_temp,2,quantile,probs=.975)
  coef_ggplot_temp$lower95 <- apply(coefs_temp,2,quantile,probs=.025)
  coef_ggplot <- rbind(coef_ggplot,coef_ggplot_temp)
}
```

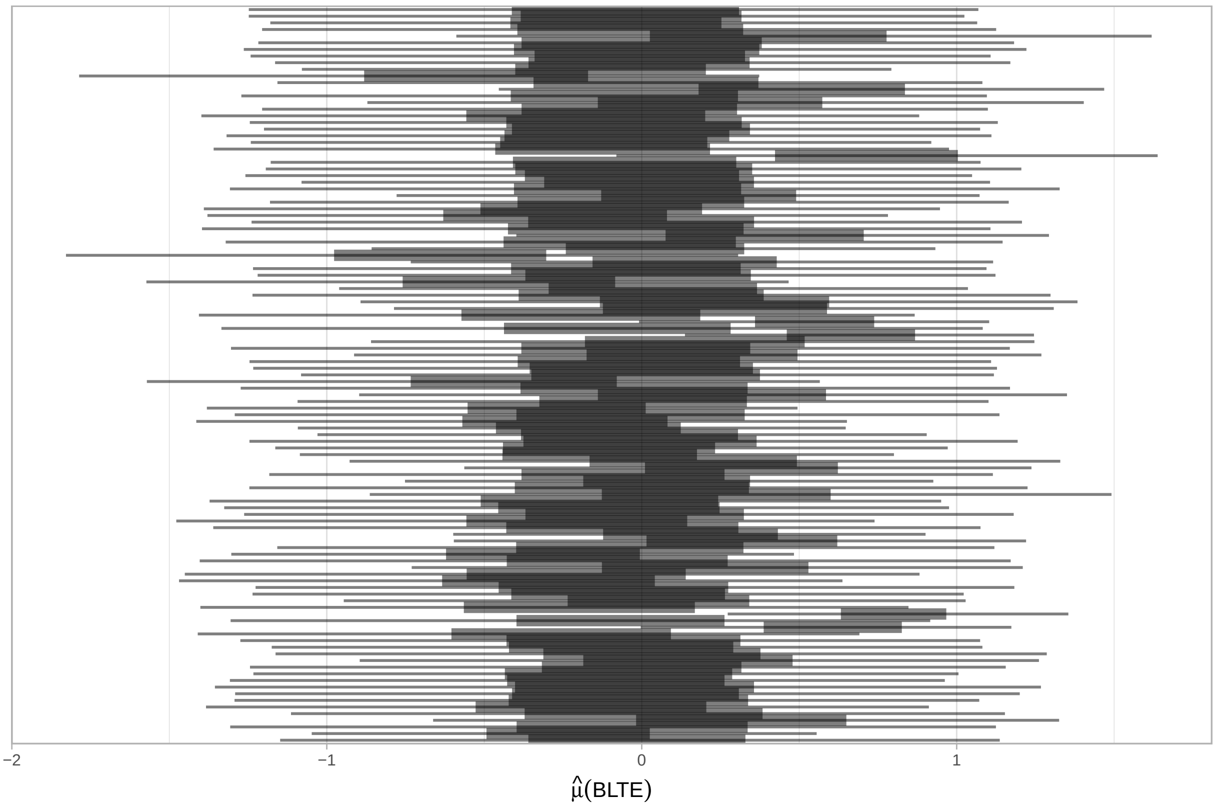
```
ggplot(coef_ggplot[coef_ggplot$wthn != "BLTE_e",]) +
  theme_light(8) +
  aes(numb, coef_mean,color=substr(wthn,1,4)) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  coord_flip() +
  labs(title="Figure 3.1: Analysis of Variance, Within",
       y=expression(hat(mu)(.)),x="",
       color = "") +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = c(.85, 0.3))
```

Figure 3.1: Analysis of Variance, Within



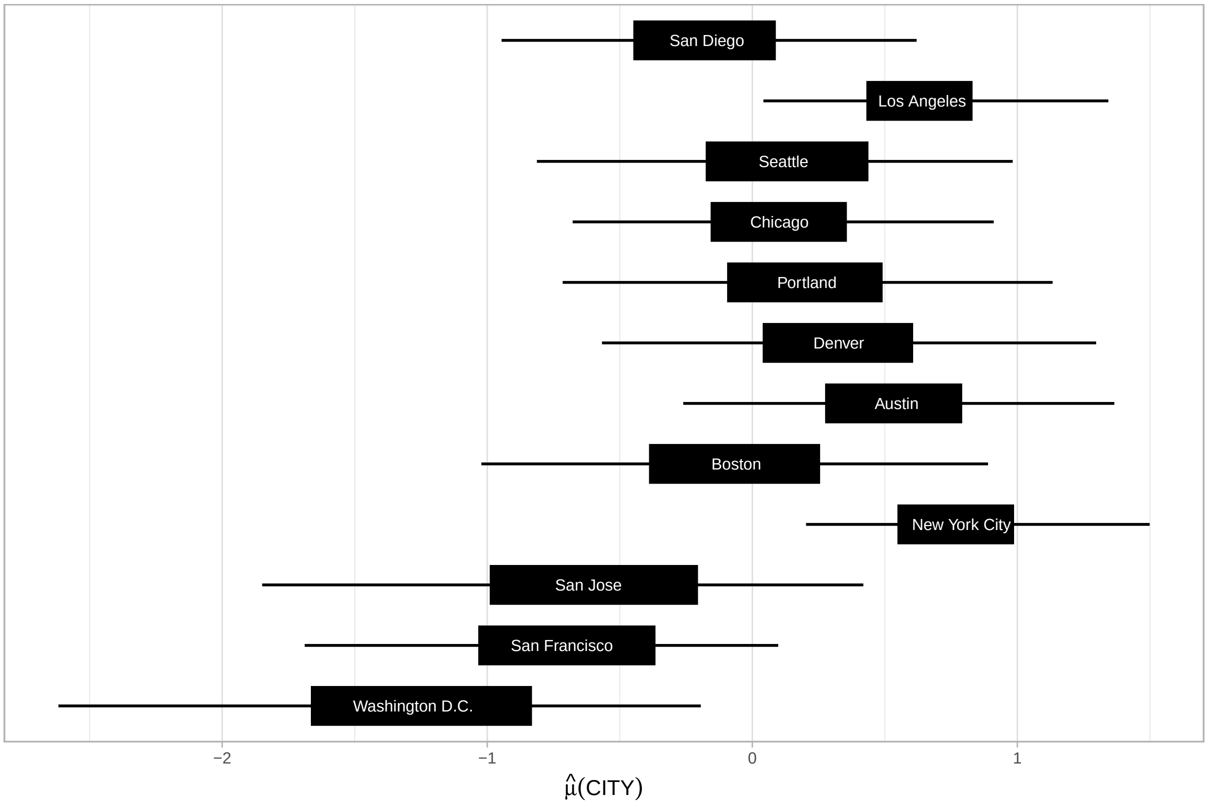

```
ggplot(coef_ggplot[coef_ggplot$wthn =="BLTE_e",]) +
  theme_light(8) +
  aes(numb, coef_mean) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2, alpha = .5) +
  geom_linerange(aes(ymin = lower95, ymax = upper95), alpha = .5) +
  coord_flip() +
  labs(title="Figure 3.2: Analysis of Variance, Within",
       y=expression(hat(mu)("BLTE")),x="",
       color = "") +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = "none")
```

Figure 3.2: Analysis of Variance, Within



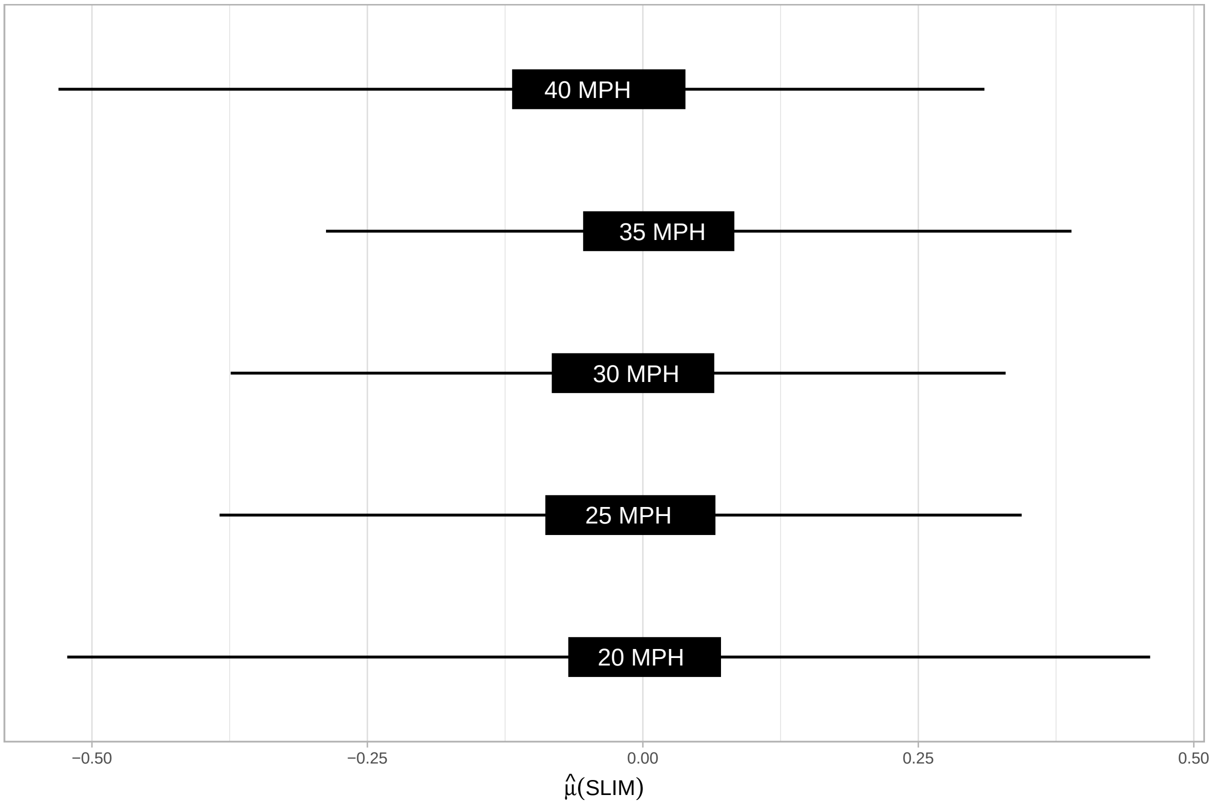

```
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "CITY_e",],
                  city_names = c("Washington D.C.","Boston","Austin",
                                 "Denver","Portland","Chicago",
                                 "Seattle","Los Angeles","San Diego",
                                 "San Francisco","San Jose",
                                 "New York City"))) +
  theme_light(8) +
  aes(numb, coef_mean) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  geom_text(aes(label = city_names),color="white",size=2) +
  coord_flip() +
  labs(title="Figure 3.3: Analysis of Variance, Within",
       y=expression(hat(mu)("CITY")),x="",
       color = "") +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = "none")
```

Figure 3.3: Analysis of Variance, Within




```
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "SLIM_e",][5:9,],
                  slim_names = c("20 MPH","25 MPH","30 MPH",
                                 "35 MPH","40 MPH"))) +
  theme_light(8) +
  aes(numb, coef_mean) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  geom_text(aes(label = slim_names),color="white",size=3) +
  coord_flip() +
  labs(title="Figure 3.4: Analysis of Variance, Within",
       y=expression(hat(mu)("SLIM")),x="",
       color = "") +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = "none")
```

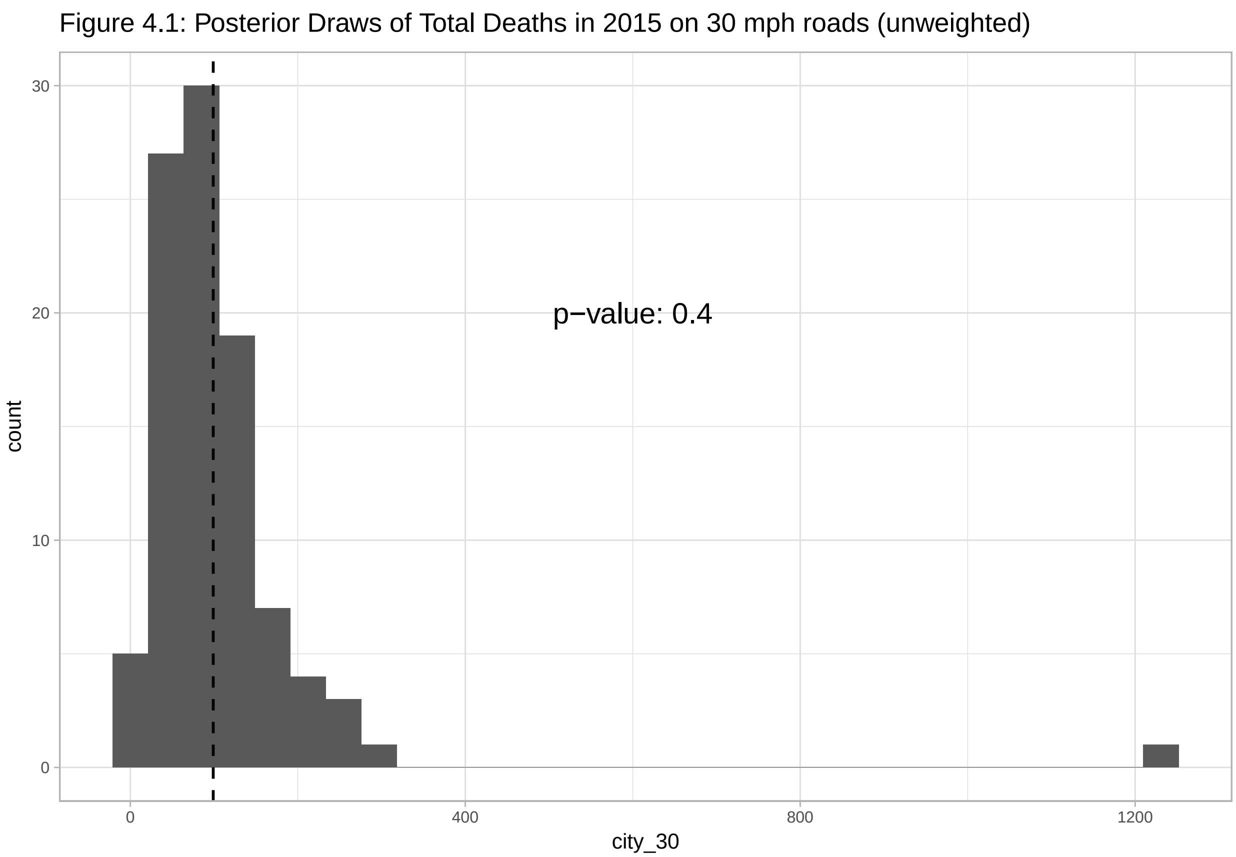
Figure 3.4: Analysis of Variance, Within



Model Checking

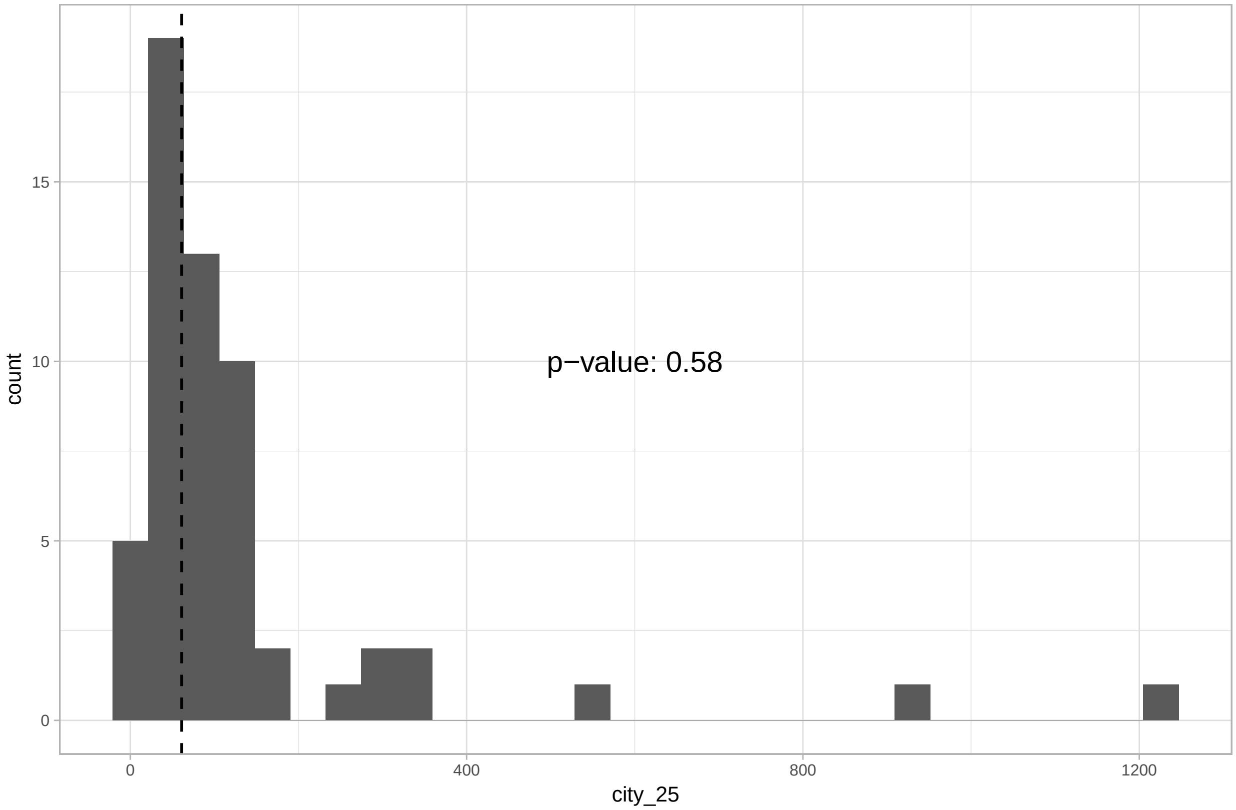
We can check our model by comparing the posterior predictions for the number of deaths on the 2015 sample roads with the actual number of deaths observed on those roads. P-values demonstrate that the number of deaths observed coincides with the model predictions. A few iterations yield unrealistically high estimates. We believe this behavior is due to the fact that large samples of the variance parameters can yield to large predictions across all strata, and a future version of this model may choose to better constrain the variance parameters.

```
pred25 <- extract(fit,c("y_pred25"))[[1]]
pred30 <- extract(fit,c("y_pred30"))[[1]]

city_30 <- apply(pred30[,stan_dat$SLIM[stan_dat$YEAR == 6] == 8],2,sum)
pvalue_30 <- paste("p-value:",round(sum(city_30 > sum(stan_dat$V1[
   stan_dat$SLIM == 8 & stan_dat$YEAR == 6]))/length(city_30),2))

qplot(city_30) +
  theme_light(8) +
  geom_vline(xintercept = sum(stan_dat$V1[stan_dat$SLIM == 8 & stan_dat$YEAR == 6]),
             linetype=2) +
  labs(title=
  "Figure 4.1: Posterior Draws of Total Deaths in 2015 on 30 mph roads (unweighted)") +
  geom_text(aes(label=pvalue_30),x=600,y=20)
```




```
city_25 <- apply(pred25[,stan_dat$SLIM[stan_dat$YEAR == 6] == 7],2,sum)
pvalue_25 <- paste("p-value:",round(sum(city_25 > sum(stan_dat$V1[
   stan_dat$SLIM == 7 & stan_dat$YEAR == 6]))/length(city_25),2))

qplot(city_25) +
  theme_light(8) +
  geom_vline(xintercept = sum(stan_dat$V1[stan_dat$SLIM == 7 & stan_dat$YEAR == 6]),
             linetype=2) +
  labs(title=
  "Figure 4.2: Posterior Draws of Total Deaths in 2015 on 25 mph roads (unweighted)") +
  geom_text(aes(label=pvalue_25),x=600,y=10)
```

Figure 4.2: Posterior Draws of Total Deaths in 2015 on 25 mph roads (unweighted)



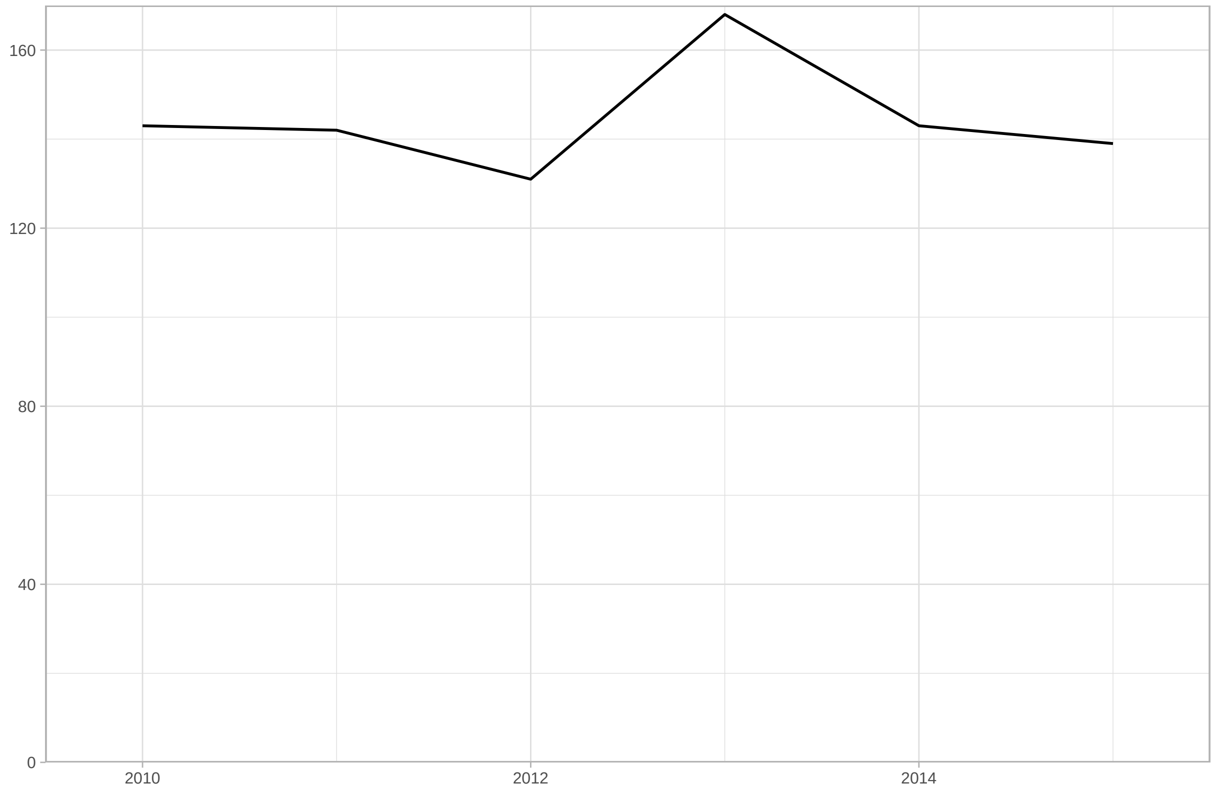
III. Policy Analysis

A before-after study would compare the number of pedestrian deaths in New York City in 2015 with the number of pedestrian deaths in New York City in 2014 or the 2010-2014 average. These types of studies are extremely popular among policy analysts because the same units are compared before and after treatment (Hauer 2005; Hauer 2015). The number of pedestrian deaths in New York City each year is displayed in Figure 5, and a before-after comparison would estimate four or six lives saved respectively. This estimate, however, fails to account for possible factors that influence the number of pedestrian deaths and changed in tandem with speed limits.

One solution would be to restrict the types of comparisons. For example, a policy analyst might choose to only compare two-lane roads with a moderate amount of traffic. An extreme case would be to construct an estimator by matching. We chose the opposite approach. We expanded the dataset beyond New York City allowing us to better learn the parameters describing roads that exist in multiple cities.

```r
ggplot(aggregate(V1 ~ CITY + YEAR,
                 stan_dat[stan_dat$CITY==12,],sum)) +
  theme_light(8) +
  aes(YEAR+2009,V1) +
  geom_line() +
  labs(x = "", y = "",
       title="Figure 5: Number of Pedestrian Deaths in NYC") +
  scale_x_continuous(limits=c(2009.5,2015.5),
                     expand = c(0, 0)) +
  scale_y_continuous(limits=c(0,170),expand = c(0,0))
```

Figure 5: Number of Pedestrian Deaths in NYC



We now calculate the effect of New York City's speed limit reduction using the posterior predictions from the model in the previous section both with and without the sampling weights from the dataset. Without the sampling weights, the estimate corresponds to the set of road regions in New York City for which there was a death in 2015. With the sampling weights, the estimate corresponds to all road regions in New York City. In either case, the benefit of lowering the speed limit is estimated to be much smaller than the before-after estimate of four or six.

```
nyc_pred25 <- pred25[,stan_dat$SLIM[stan_dat$YEAR==6] == 6 &
                      stan_dat$CITY[stan_dat$YEAR==6] == 12]
nyc_pred30 <- pred30[,stan_dat$SLIM[stan_dat$YEAR==6] == 6 &
                      stan_dat$CITY[stan_dat$YEAR==6] == 12]
WGHT <- stan_dat$WGHT[which(stan_dat$SLIM[stan_dat$YEAR==6] == 6 &
        stan_dat$CITY[stan_dat$YEAR==6] == 12)]

kable(rbind(
  round(quantile(apply(nyc_pred25 - nyc_pred30, 2, mean)),2),
  round(quantile(apply(nyc_pred25 - nyc_pred30, 2, mean) * WGHT),2)))
```

0%	25%	50%	75%	100%
-0.09	0	0	0.01	0.10
-0.04	0	0	0.00	0.03

IV. Appendix

Data Sources

Our primary dataset is the Fatality Analysis Reporting System (FARS) collected by the National Highway Traffic Safety Administration. FARS contains a census of vehicle related deaths in the United States with a large number of covariates detailing the road, vehicle and persons involved. Road refers to the general area in which the death took place. This is generally a road segment or an intersection.

We supplement the FARS data with a few other sources. These sources are not included in the reference but their websites have been linked to here. We obtain probability weights for each FARS observation using the GES. Missing New York City speed limit data is obtained from NYC DOT using the shapefiles of speed limits. The pedestrian population on each road was calculated as suggested by the Census Bureau using the TPP and PDB. Average annual daily traffic was estimated from HPMS. The TFFC variable is discretized by the number of vehciles per second.

V. References

Fatality Analysis Reporting System (FARS) 2010-2015 Final File, 2015 Annual Report File (ARF)

Gelman, A. & Hill, J. 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. 2013. Bayesian Data Analysis. CRC Press, Boca Raton, FL, USA.

Hauer, E. The Art of Regression Modeling in Road Safety. Springer International Publishing, Cham, Switzerland, 2015.

Hauer, E. "Cause and effect in observational cross-section studies on road safety." Unpublished manuscript, 2005.

Goodwin A, Thomas L, Kirley B, Hall W, O'Brien N, Hill K. Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices, Eighth Edition. Report no. DOT HS-812-202. National Highway Traffic Safety Administration, Washington, DC; 2015.

Leaf, W.A., & Preusser, D.F. (1999). Literature review on vehicle travel speeds and pedestrian injuries (Report No. DOT HS 809 021). Retrieved from the NHTSA website: www.nhtsa.dot.gov/people/injury/research/pub/HS809012.html

Tingvall C. and Haworth N., Vision Zero: an ethical approach to safety and mobility Proceedings of the 6th ITE International Conference on Road Safety and Traffic Enforcement; Beyond 2000, Melbourne, 6-7 September (1999)